PB87-115325

Finite Expansion Method for the Calculation and Interpretation of Molecular Electrostatic Potentials

(U.S.) Health Effects Research Lab.
Research Triangle Park, NC

1986

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completi

1. REPORT NO.	2.	3.
EPA/600/J-86/171		PB87-115325

4. TITLE AND SUBTITLE
A Finite Expansion Method for the Calculation and Interpretation of Molecular Electrostatic Potentials (Journal Version)

5. REPORT DATE
1986

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Rabinowitz, Namboodri, and Weinstein

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
EPA, CMB, Research Triangle Park, NC
Mt. Sinai School of Medicine, City University of NY, NY, NY

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.

12. SPONSORING AGENCY NAME AND ADDRESS
Health Effects Research Laboratory - RTP, NC
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

13. TYPE OF REPORT AND PERIOD COVERED

14. SPONSORING AGENCY CODE
EPA/600/11

15. SUPPLEMENTARY NOTES
REF: International J. of Quantum Chem., XXIX:1697-1704, June 1986

16. ABSTRACT

Because it is useful to have the molecular electrostatic potential as an element in a complex scheme to assess the toxicity of large molecules, efficient and reliable methods are needed for the calculation and characterization of these potentials. A multicenter multipole expansion of the molecular electron charge density calculated with a limited Gaussian basis set is shown here to have only a finite number of nonzero terms from which the molecular electrostatic potential can be calculated. The discrete contributions to the electrostatic potentials from the terms of this expansion provide a physically meaningful decomposition of the potential and a means for its characterization. With pyrrole as an example, the electrostatic potential calculated from this finite expansion of the electron density is compared to that obtained from exact calculations from the same wave function. Good agreement is obtained at distances greater than 1.5 A from any atom in the molecule. In contrast, rearrangement of the terms into an expansion corresponding only to Mulliken atomic charges and dipoles yields a decomposition that produces electrostatic potentials which agree less with the exact potential. This descrepancy is attributable to the neglect of terms due to higher moments.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	10
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

EPA/600/J-86/171
JOURNAL ARTICLE

INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY, VOL. XXIX, 1697–1704 (1986)

A Finite Expansion Method for the Calculation and Interpretation of Molecular Electrostatic Potentials

JAMES R. RABINOWITZ
Environmental Protection Agency, Research Triangle Park, North Carolina 27711, USA

KRISHNAN NAMBOODIRI and HAREL WEINSTEIN
Departments of Pharmacology and of Physiology and Biophysics, Mount Sinai School of Medicine of the City University of New York, New York, New York 10029, USA

Abstract

Because it is useful to have the molecular electrostatic potential as an element in a complex scheme to assess the toxicity of large molecules, efficient and reliable methods are needed for the calculation and characterization of these potentials. A multicenter multipole expansion of the molecular electron charge density calculated with a limited Gaussian basis set is shown here to have only a finite number of nonzero terms from which the molecular electrostatic potential can be calculated. The discrete contributions to the electrostatic potentials from the terms of this expansion provide a physically meaningful decomposition of the potential and a means for its characterization. With pyrrole as an example, the electrostatic potential calculated from this finite expansion of the electron density is compared to that obtained from exact calculations from the same wave function. Good agreement is obtained at distances greater than 1.5 A from any atom in the molecule. In contrast, rearrangement of the terms into an expansion corresponding only to Mulliken atomic charges and dipoles yields a decomposition that produces electrostatic potentials which agree less well with the exact potential. This discrepancy is attributable to the neglect of terms due to higher moments.

Introduction

The relationships between chemical structure, physicochemical properties, and activity in a biological system provide the rationale for the development and evaluation of reliable computational methods for assessing the possible adverse health effects of chemicals [1–3]. When the chemicals are not available for bioassay, or in situations where the relevant bioassays are incomplete, such theoretical approaches have considerable practical significance for the development of regulatory, testing and experimental research strategies [1]; their advantages have recently been discussed in detail [2]. In particular, because the electrostatic interaction has been shown to be a critical component of the intermolecular interactions in many biological processes [4,5], molecular electrostatic potentials have been found to provide useful insight into the differential biological activity within families of chemicals [3,4].

© 1986 John Wiley & Sons, Inc.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

Due to the large variety of processes that may result in an adverse health effect of a xenobiotic [2], it is probable that any useful scheme to assess toxicity even within a family of chemicals will require the analysis of several physicochemical properties. To include the molecular electrostatic potential generated by the chemicals of interest as an element in this type of analytic and predictive scheme, efficient and reliable methods are needed for its calculation and its characterization in terms of a few quantitative characteristics. A physically meaningful decomposition of the molecular charge distribution that can lead to such a characterization of the electrostatic potential is described here. It is based on a method for the decomposition of a molecular charge distribution calculated from a wavefunction expanded in a limited gaussian basis set. This decomposition has a finite number of nonzero terms [6]. Its use for the calculation of the electrostatic potential is compared to that with the more traditional atomic multipole expansion. The characteristics it provides are discussed with reference to the molecular electrostatic potential calculated exactly from the same wave function.

Decomposition of the Molecular Charge Distribution

For any molecular charge distribution $\rho(\bar{r})$ described by a limited gaussian basis set, there is a multicenter, multipole expansion that has only a finite number of terms that can all be specified [6–8]. The charge distribution is written in terms of the gaussian basis as

$$\rho(\bar{r}) = \sum_{i,j} c_{ij} g_i(\alpha_i, \bar{r}_i) g_j(\alpha_j, \bar{r}_j) \quad (1)$$

where $\rho(\bar{r})$ is the molecular charge distribution and g_i and g_j are standard cartesian gaussians with corresponding quantum numbers n,l,m.

Since the product of any cartesian gaussians is a sum of individual cartesian gaussian about a single center, the charge distribution can also be written as a simple sum of single gaussians defined about the point N:

$$\rho(\bar{r}) = \sum_{N} \rho_N(\bar{r}). \quad (2)$$

If g_i and g_j were both centered on the same atom, then ρ_N is just a single gaussian about that atom with $l = l_i + l_j$. If they were defined about different atoms, then $\bar{R}_N$, the position of the point N, is

$$\bar{R}_N = \frac{\alpha_i \bar{R}_i + \alpha_j \bar{R}_j}{\alpha_i + \alpha_j}, \quad (3)$$

and the sum includes values of l up to $l_i + l_j$.

The electrostatic potential $V(\bar{r})$ generated by this charge distribution is

$$V(\bar{r}) = \int \frac{\rho(\bar{r}')}{|\bar{r} - \bar{r}'|} d\bar{r}' = \sum_{N} \int \frac{\rho_N(\bar{r}')}{|\bar{r} - \bar{r}'|} d\bar{r}' \quad (4)$$

where the sum N is over all the expansion centers. According to [9]:

$$\int \frac{\rho_N(\vec{r}')}{|\vec{r}-\vec{r}'|}\,d\vec{r}' = 4\pi \sum_{l,m} \frac{1}{2l+1} \int Y^*_{l,m}(\theta',\varphi')\vec{r}'^{\,l}\rho_N(\vec{r}')\,d\vec{r}' \cdot Y_{l,m}(\theta,\varphi)$$

$$\cdot \frac{1}{\vec{r}^{\,l+1}} = 4\pi \sum_{l,m} \frac{1}{2l+1} Q_{l,m} \cdot Y_{l,m}(\theta,\varphi) \cdot \frac{1}{r^{l+1}} \tag{5}$$

for any $\vec{r} > \vec{r}'$. To be consistent with the form of the wave function derived from the GAUSSIAN 80 program package [10], the Y_{lm} functions are chosen as the real spherical harmonics.

Since each $\rho_N(\vec{r}')$ contains only a finite number of terms with no value of l greater than $l_i + l_j$, and $Y_{lm}(\theta,\varphi)$ is defined in the same coordinate system as ρ_N, the expansion in Eq. (5) only contains a few terms. The physical meaning of these terms is reflected in the coefficients $Q_{l,m}$: $Q_{0,0}$ is the total charge in ρ_N multiplied by $(1/\sqrt{4\pi})$; the $Q_{1,m}$ terms are the cartesian dipole moments multiplied by $\sqrt{3/4\pi}$; and the $Q_{2,m}$ terms are the 5 quadruple contributions of the form $(2z^2 - x^2 - y^2)$; $(x^2 - y^2)$, xy, xz, and yz, respectively. For a basis set including only s and p functions the expansion is complete with these terms.

The contribution (w_i) to the total molecular electrostatic potential due to each of these terms is

$$w_1 = \frac{\sqrt{4\pi}}{r} Q_{0,0}, \tag{6a}$$

$$w_2 = \frac{\sqrt{4\pi/3}}{r^2} (Q_{1,1} \sin\theta\cos\varphi + Q_{1,-1} \sin\theta\sin\varphi + Q_{1,0}\cos\theta), \tag{6b}$$

$$w_3 = \frac{\sqrt{\pi/5}}{r^3} [Q_{2,0}(3\cos^2\theta - 1) + \sqrt{3}\,Q_{2,-2}\sin^2\theta(\cos^2\varphi - \sin^2\varphi)$$
$$+ 2\sqrt{3}(Q_{2,2}\sin^2\theta\cos\varphi\sin\varphi + Q_{2,1}\cos\theta\sin\theta\cos\varphi$$
$$+ Q_{2,-1}\cos\theta\sin\theta\sin\varphi)]. \tag{6c}$$

The multicenter, multipole expansion resulting from Eq. (2) can also be contracted to the standard description in terms of atomic charges (i.e., the Mulliken charges) and multipoles by assigning half the charge in two center charge distributions to each atomic center and using standard multipole translation equations [11]. The translation introduces an infinite number of nonzero terms. Schemes that also include midpoint multipole expansions [5] can also be reproduced by using the translation equations. They will also have an infinite number of nonzero terms, since the natural center (Eq. 3) for most of the two center charge distributions is not the atomic midpoint. In fact, for atoms of different types the midpoint is not the natural expansion center for any two center charge distributions.

Results and Discussion

The use of the decomposition of the charge distribution in the calculation of electrostatic potentials is illustrated for pyrrole. Equations (2) and (5) were applied to a molecular charge distribution calculated for pyrrole with the STO-3G basis set and the GAUSSIAN 80 system of programs [10]. The resulting electrostatic potentials were compared with exact calculations [4].

The finite expansion of the charge distribution of pyrrole according to the method presented above yields 92 centers at which the charge distributions have a charge of absolute magnitude greater than 0.01 electron units. Each pair of gaussians defines a charge distribution at one center. Because many pairs have the same expansion point (Eq. 3) they are grouped to represent the charge distribution at that point.

The ten atoms of pyrrole are included among 15 of the 92 centers at which the absolute magnitude of the charge is greater than 0.3 electron units. The other five points are located between nearest neighbor *C*—*C* and *C*—*N* pairs. This location of charge distributions of large magnitude is a result of the properties of an STO-*n*G basis set [12]. The position $\bar{R}$ of these additional points is defined by

$$\bar{R} = \frac{\bar{R}_i + t\bar{R}_j}{1 + t}, \tag{7}$$

where t is the ratio of the squares of the appropriate Slater screening parameters and R is the exact midpoint of the bond only for identical atoms. Fifty-one two-center products of gaussians contribute to the charge distribution at each of these points. Because they result from several gaussian pairs that include functions other than *s*-type, the charge distributions at the five main bond points have both dipole and quadrupole moments in addition to their charge component. Their contribution to the potential is therefore likely to be more important than the one center hydrogen charge distributions which do not have higher moments.

There are 29 points with charges of absolute magnitude greater than 0.1 but less than 0.3 electron units, all of them positioned along the chemical bonds in the molecule. Of the additional 48 points with charges greater than 0.01 electron units, 31 points are contributed by gaussians on nonnearest neighbor atoms.

The electrostatic potential as a function of position along the symmetry axis of the molecule (the direction of the *N*—*H* bond, which is placed along the *Z* axis) in a plane positioned 2.9 bohr above the molecule is shown in Figure 1. Near the molecule, the contribution to the potential due to point dipoles (Eq. 6b) is opposite in sign and nearly equal in magnitude to that due to point charges (Eq. 6a). Consequently, the quadrupole terms (Eq. 6c) becomes a dominant contribution to the potential in this region. The potential decays to zero as $1/r^2$ at large values of Z.

As previously described, the large number of individual charge distributions can be translated to give a set of atomic moments. There will be an infinite number of multipoles at each atom. Figure 2 shows the potential calculated from such an expansion truncated after its second term, i.e., including only the potential due to atomic charges (defined as Mulliken charges) and atomic dipoles. Figure 3 compares these potentials to that computed exactly from the same wavefunction. The potential

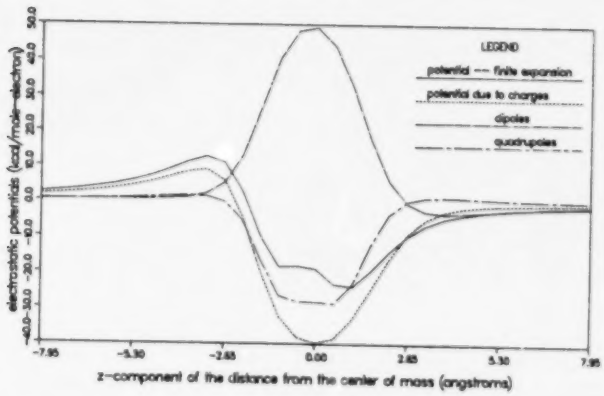

Figure 1. The molecular electrostatic potential of pyrrole generated by the terms of the finite expansion of the charge distribution calculated with the STO-3G basis set. Values are calculated along the symmetry (Z) axis of the molecule in the direction of the N—H bond in a plane positioned 2.9 bohr above the molecule. The contributions from each type of multipole are shown separately.

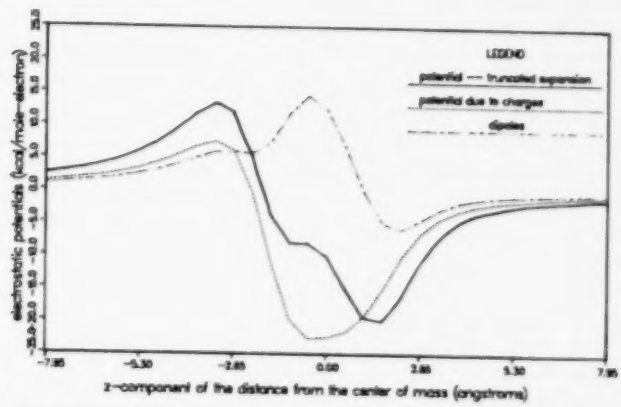

Figure 2. The molecular electrostatic potential of pyrrole generated by an atomic expansion of the electron density truncated to include only the charge and dipole terms. Other details as in Figure 1.

calculated for the finite expansion is in better numerical agreement with the exact one at all distances and reproduces the shape of the curve near the minimum, although the magnitude of the minimum differs by about 20%. At 1.5 A from the minimum the potential calculated from the finite expansion method agrees with the exact one to an accuracy of a few percent. In contrast, the truncated expansion displaces the minimum by 1 A and fails to give as good a numerical agreement as the finite expansion near the molecule, i.e., within the potential well. All three calculated potentials agree at distances greater than 2.5 A from any atom in the pyrrole molecule and approach zero at large distances with a $1/r^2$ dipole-like dependence.

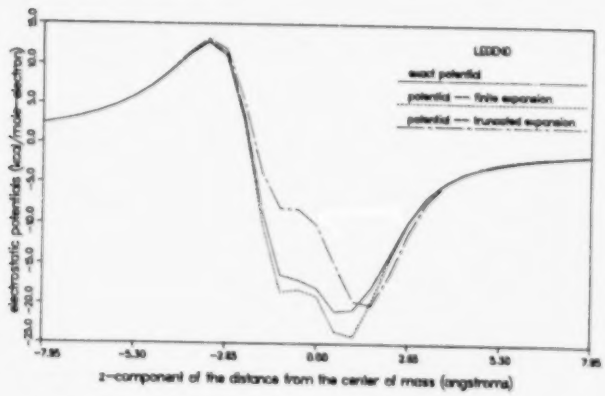
LEGEND
exact potential
potential — finite expansion
potential — truncated expansion
electrostatic potentials (kcal/mole-electron)
z-component of the distance from the center of mass (angstroms)

Figure 3. Comparison of molecular electrostatic potentials of pyrrole calculated with an exact method and from the finite and the truncated multipole expansions. See Figure 1 and 2 for details.

Figure 4 shows the variation in the potential (calculated in the three ways described above) in the plane of the molecule but in a direction perpendicular to the molecular dipole moment. All three potentials approach zero with a $1/r^3$ quadrupole-like dependence for large r values. The potential generated by the finite expansion method is within 7% of the exact potential at 1.66 A from the nearest atom and is essentially identical at all points significantly beyond that distance. In contrast, there is at least a 10% error at all distances for the truncated expansion. For large distances the error is greater than 15%. For both expansions the electrostatic potential due to either charge or dipole terms approaches zero as $1/r^3$. The failure of the truncated expansion at long distances results from the omission of atomic quadrupoles, a term that also approaches

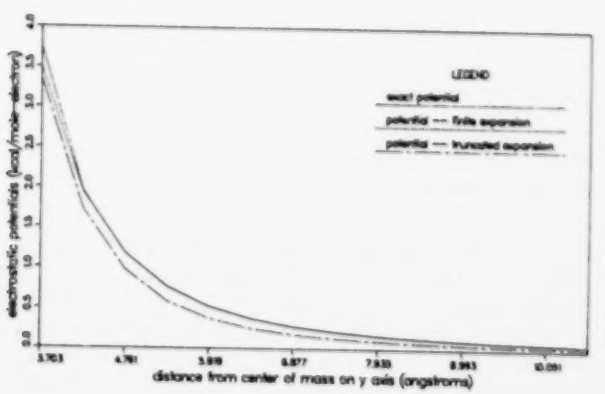

Figure 4. Comparison of molecular electrostatic potentials of pyrrole in a direction perpendicular to the molecular plane, along an axis through the center of mass. Values are calculated with an exact method and from the finite and the truncated multipole expansions.

zero as $1/r^3$. In other directions, not perpendicular to the molecular dipole moment, this omission is not important for large values of r. This would also be the case for charged molecules for which the potential decays as $1/r$.

The finite expansion technique retains all the terms. The number of these terms depends on the highest angular momentum in the basis set used for the wave function. If orbitals higher than p are used, the expansion will contain multipole moments higher than quadrupole. The discrepancies between the potential calculated by this technique and the exact potential are due to the assumption, implicit in Eq. (5) that the point at which the potential is calculated resides outside the charge distribution; but the gaussian wave functions are infinite in extent. The discrepancy results from the molecular charge farther from the atomic nuclei than the point at which the potential is calculated. This expansion, and any other expansion similar to Eq. (5), is therefore not useful near the atomic centers.

Conclusion

There is a multicenter, multipole expansion that has only a finite number of terms for a limited basis set gaussian wave function.

Using the charge distribution calculated for pyrrole with an STO-3G basis set, and selecting appropriate expansion centers, we showed that in the potential calculated from this expansion only the first three terms are nonzero. These are due to charge, dipole and quadrupole contributions. In this basis set there is an expansion point between nearest neighbor heavy atoms that contributes as much to the potential as the hydrogens; it is the bond midpoint only for identical atoms. Good agreement is obtained between the potential calculated from this expansion and the exact potential at distances greater than 1.5 A away from any atom.

Constructing the expansion points to give Mulliken atomic charges and corresponding dipoles fails to yield as good an agreement as the original expansion in regions close to the atomic centers, and never yields agreement with the exactly calculated potential in directions perpendicular to the molecular dipole. This result is attributable to the neglect of quadrupole terms and illustrates the point that any expansion technique must preserve the molecular quadrupole moments if it is to give a reliable electrostatic potential in all directions.

Acknowledgment

This work was supported in part by the US Environmental Protection Agency under cooperative research agreement CR 811006. H. W. is a recipient of a Research Scientist Development Award (DA-00060) from the National Institute on Drug Abuse. A generous grant of computer time from the University Computing Center of the City University of New York is gratefully acknowledged. The research described has been reviewed by the Health Effects Research Laboratory, U.S. Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflects the views and policies of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Bibliography

[1] J. R. Rabinowitz, Ed. *EPA Report: "Research Strategies for the Application of the Techniques of Computational Biological Chemistry to Environmental Problem"* (1985).
[2] H. Weinstein, J. Rabinowitz, M. N. Liebman, and R. Osman, Environm. Health Perspect. **61**, 147, 1985.
[3] H. Weinstein, K. Namboodiri, R. Osman, M. N. Liebman, and J. R. Rabinowitz, in *QSAR in Toxicology and Xenobiochemistry*, M. Tichy, Ed. (Elsevier Science Publishers, Amsterdam, 1985), p. 451.
[4] H. Weinstein, R. Osman, J. P. Green, and S. Topiol, in *Chemical Applications of Atomic and Molecular Electrostatic Potentials*. P. Politzer and D. G. Truhla, Eds. (Plenum, New York, 1981), pp. 309–334.
[5] A. Pullman and B. Pullman, Ann. Rev. Biophys. **7**, 505 (1975).
[6] A. J. Stone, Chem. Phys. Letters **83**, 233 (1981).
[7] D. Martin and G. G. Hall, Theoret. Chima Acta. **59**, 281, 1981.
[8] L. L. Shipman, Chem. Phys. Letters **31**, 361 (1975).
[9] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), p. 99.
[10] J. S. Binkley, R. A. Whiteside, R. Krishnan, R. Seeger, D. J. DeFrees, H. B. Schlegel, S. Topiol, L. R. Kahn, and J. A. Pople, *GAUSSIAN 80, IBM version*, unpublished.
[11] J. R. Rabinowitz, T. J. Swissks, and R. Rein, Int. J. Quantum Chem. **56**, 353, 1972.
[12] W. J. Hehre, R. F. Stewart and J. A. Pople, J. Chem Phys. **51**, 2657 (1969).

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.